Input vector: $\mathbf{x} \in \mathbb{R}^D$

Transformation matrix: $\mathbf{A} = \begin{bmatrix} | & & | \\ \boldsymbol{\alpha_1} & \dots & \boldsymbol{\alpha_M} \\ | & & | \end{bmatrix} \in \Re^{D\times M}$

Finding 1st PCA dimension $\boldsymbol{\alpha_1} \in \mathbb{R}^D$

We would like to find $\boldsymbol{\alpha_1} \in \mathbb{R}^D$ that maximizes the variance:

$$Var\{\boldsymbol{\alpha_1}^T \mathbf{x}\} = \boldsymbol{\alpha_1}^T \boldsymbol{\Sigma} \boldsymbol{\alpha_1}$$

where $\boldsymbol{\Sigma}$ is the covariance of the data $\{\mathbf{x_1}, \dots, \mathbf{x_N}\}$.

Constrained optimization problem:

$$\max \boldsymbol{\alpha_1}^T \boldsymbol{\Sigma} \boldsymbol{\alpha_1} \ , \ \ s.t. \ \ \boldsymbol{\alpha_1}^T \boldsymbol{\alpha_1} = 1$$

Lagrange optimization:

$$L = \boldsymbol{\alpha_1}^T \boldsymbol{\Sigma} \boldsymbol{\alpha_1} - \lambda(\boldsymbol{\alpha_1}^T \boldsymbol{\alpha_1} - 1)$$
$$\Rightarrow \frac{\theta L}{\theta \boldsymbol{\alpha_1}} = 0 \Rightarrow \boldsymbol{\Sigma} \boldsymbol{\alpha_1} - \lambda \boldsymbol{\alpha_1} = 0 \Rightarrow \lambda \boldsymbol{\alpha_1} = \boldsymbol{\Sigma} \boldsymbol{\alpha_1}$$

This is the eigenvector equation! We choose the eigenvector with the largest eigenvalue.

Finding 2nd PCA dimension $\boldsymbol{\alpha_2} \in \mathbb{R}^D$

We would like to find $\boldsymbol{\alpha_2} \in \mathbb{R}^D$ that maximizes the variance $Var\{\boldsymbol{\alpha_2}^T \mathbf{x}\}$, so that $\boldsymbol{\alpha_2}$ is *orthogonal* to $\boldsymbol{\alpha_1}$.

Constrained optimization problem:

$$\max \boldsymbol{\alpha_2}^T \boldsymbol{\Sigma} \boldsymbol{\alpha_2} \ , \ \ s.t. \ \ \boldsymbol{\alpha_2}^T \boldsymbol{\alpha_2} = 1 \ \ and \ \ \boldsymbol{\alpha_2}^T \boldsymbol{\alpha_1} = 0$$

Lagrange optimization:

$$L = \boldsymbol{\alpha_2}^T \boldsymbol{\Sigma} \boldsymbol{\alpha_2} - \lambda(\boldsymbol{\alpha_2}^T \boldsymbol{\alpha_2} - 1) - \phi \boldsymbol{\alpha_2}^T \boldsymbol{\alpha_1}$$
$$\Rightarrow \frac{\theta L}{\theta \boldsymbol{\alpha_2}} = 0 \Rightarrow \boldsymbol{\Sigma} \boldsymbol{\alpha_2} - \lambda \boldsymbol{\alpha_2} - \phi \boldsymbol{\alpha_1} = 0$$

If we left multiply $\boldsymbol{\alpha_1}$ in the above expression:

$$\boldsymbol{\alpha_1}^T \boldsymbol{\Sigma} \boldsymbol{\alpha_2} - \lambda \boldsymbol{\alpha_1}^T \boldsymbol{\alpha_2} - \phi \boldsymbol{\alpha_2}^T \boldsymbol{\alpha_1} = 0 \Rightarrow \phi = 0$$

When $\phi = 0$, we get $\lambda \boldsymbol{\alpha_2} = \boldsymbol{\Sigma} \boldsymbol{\alpha_2}$. This corresponds to another eigenvalue equation and we choose the eigenvector with the second largest eigenvalue.